ON THE MOMENTS OF RATIO-BASED ESTIMATORS IN JOIN POINT ESTIMATION

Walter W. Piegorsch
Biometrics Unit, Cornell University, Ithaca, New York 14853

BU-757-M*   Revised December 1981

Abstract

The maximum likelihood estimator (MLE) in bilinear segmented regression can take on the form of a ratio of correlated random variables. Calculation of the moments of the sampling distribution of this ratio is attempted, with emphasis on the mean and variance. A modified estimate based on the MLE is considered, and its moments are also explored.

---

* In the Biometrics Unit Series, Cornell University, Ithaca, New York 14853.

ON THE MOMENTS OF RATIO-BASED ESTIMATORS IN JOIN POINT ESTIMATION

Walter W. Piegorsch

In experimental situations concerning both the physical and social sciences, the functional relationship of a dependent random variable upon another (independent) variable is often not of a simple form – e.g., polynomial – but of a more complicated form: a spline function. The simplest of these is the bilinear spline, where two lines intersect at a join point:

$$E(y_i) = \begin{cases} \alpha_1 + \beta_1 x_i & i = 1, \cdots, \tau \\ \alpha_2 + \beta_2 x_i & i = \tau+1, \cdots, n \end{cases} \quad . \tag{1}$$

The literature on the estimation of the abscissa of the intersection, $J = (\alpha_1 - \alpha_2)/(\beta_2 - \beta_1)$, or on the estimation of the index value separating the two segments, τ, has grown over the years since W.R. Blischke first derived an algorithmic procedure to produce least squares estimators (Blischke, 1961). A variety of methods for this estimation have appeared, some seeming perhaps more "ad hoc" than others, ranging from the classical principle of maximum likelihood – which, when a Gaussian error structure is supposed, mimics the least squares procedure – to more recent decision theoretic approaches (see Shabon, 1980). One critical distinction many of these procedures have made is to treat separately the simplified case of τ known, i.e., where the location, in terms of the x_i's, of the two segments is known; the problem, however, still remains to estimate J. In such a case, both Blischke and D.E. Robison (1964, in a more general result on polynomial splines), showed that the least squares/maxiumum likelihood estimate is simply the intersection of the two sample regressions, a ratio of the form

$$\hat{J} = (\hat{\alpha}_1 - \hat{\alpha}_2)/(\hat{\beta}_2 - \hat{\beta}_1) \quad , \tag{2}$$

where $\hat{\alpha}_1$ and $\hat{\beta}_1$ are the least squares intercept and slope estimates, respectively, for the first τ data pairs, and $\hat{\alpha}_2$ and $\hat{\beta}_2$ are those for the next $n-\tau$ data pairs (when the estimate falls outside of the interval $[x_\tau, x_{\tau+1}]$, the endpoint which maximizes the likelihood function is taken as the join abscissa). In the less restrictive case of τ unknown, this procedure is simply applied over all τ, and that $\hat{J}_\tau$ which maximizes the likelihood is chosen as $\hat{J}$. One obvious question that occurs is, what other optimal properties does this estimator possess? We will consider the moments of $\hat{J}$, starting with $E(\hat{J})$ in the simple case τ known; the results prove to be quite interesting indeed.

With τ known, we consider the structure of (2) where, under the usual assumptions, $y_i \sim$ indep. $N[E(y_i), \sigma^2]$, we see $\hat{\alpha}_1 - \hat{\alpha}_2 \sim N(\alpha_1 - \alpha_2, \nu^2)$ for

$$\nu^2 = \sigma^2\left\{\frac{\Sigma_{i=1}^{\tau} x_i^2}{\tau\Sigma_{i=1}^{\tau}(x_i - \bar{x}_1)^2} + \frac{\Sigma_{i=\tau+1}^{n} x_i^2}{(n-\tau)\Sigma_{i=\tau+1}^{n}(x_i - \bar{x}_2)^2}\right\} \quad , \tag{3}$$

and

$$\bar{x}_1 = \Sigma_{i=1}^{\tau}(x_i/\tau) \ , \qquad \bar{x}_2 = \Sigma_{i=\tau+1}^{n}[x_i/(n-\tau)] \quad .$$

Also, $\hat{\beta}_2 - \hat{\beta}_1 \sim N(\beta_2 - \beta_1, \eta^2)$ for

$$\eta^2 = \sigma^2\left\{\frac{1}{\Sigma_{i=1}^{\tau}(x_i - \bar{x}_1)^2} + \frac{1}{\Sigma_{i=\tau+1}^{n}(x_i - \bar{x}_2)^2}\right\} \quad , \tag{4}$$

and $\mathrm{cov}(\hat{\alpha}_j, \hat{\beta}_j) = -\bar{x}_j\sigma^2/\Sigma_i(x_i - \bar{x}_j)^2$ $(j=1, 2)$ (DeGroot, 1975). Thus $\hat{J}$ is a ratio of two correlated Gaussian random variables. D.V. Hinkley (1969) has derived the density function of such a random variable, but its form – which we consider later – is complicated, and instead of using it to calculate $E(\hat{J})$ directly we will use a different approach. Writing $\hat{\alpha}_j$ in the familiar form

$$\hat{\alpha}_j = \bar{y}_j - \hat{\beta}_j\bar{x}_j \qquad (j=1, 2) \quad , \tag{5}$$

where

$$\bar{y}_1 = \sum_{i=1}^{\tau} y_i/\tau \qquad \text{and} \qquad \bar{y}_2 = \sum_{i=\tau+1}^{n} y_i/(n-\tau) \quad , \tag{6}$$

gives us a reexpression of (2) as

$$\hat{J} = \frac{\bar{y}_1 - \hat{\beta}_1\bar{x}_1 - \bar{y}_2 + \hat{\beta}_2\bar{x}_2}{\hat{\beta}_2 - \hat{\beta}_1} = \frac{\bar{y}_1 - \bar{y}_2}{\hat{\beta}_2 - \hat{\beta}_1} + \frac{\hat{\beta}_2\bar{x}_2 - \hat{\beta}_1\bar{x}_1}{\hat{\beta}_2 - \hat{\beta}_1} \quad . \tag{7}$$

Then

$$\begin{aligned} E(\hat{J}) &= E\left(\frac{\bar{y}_1 - \bar{y}_2}{\hat{\beta}_2 - \hat{\beta}_1}\right) + E\left(\frac{\hat{\beta}_2\bar{x}_2 - \hat{\beta}_1\bar{x}_1}{\hat{\beta}_2 - \hat{\beta}_1}\right) \\ &= E(\bar{y}_1 - \bar{y}_2)E[(\hat{\beta}_2 - \hat{\beta}_1)^{-1}] + \mathrm{Cov}[\bar{y}_1 - \bar{y}_2, (\hat{\beta}_2 - \hat{\beta}_1)^{-1}] \\ &\quad + E(\hat{\beta}_2\bar{x}_2 - \hat{\beta}_1\bar{x}_1)E[(\hat{\beta}_2 - \hat{\beta}_1)^{-1}] + \mathrm{Cov}[\hat{\beta}_2\bar{x}_2 - \hat{\beta}_1\bar{x}_1, (\hat{\beta}_2 - \hat{\beta}_1)^{-1}] \quad . \end{aligned}$$

In two terms of this expression the value of $E[(\hat{\beta}_2 - \hat{\beta}_1)^{-1}]$ is critical. To evaluate it consider, in general $E(1/V) = E[f(V)]$. By the definition of expectation, this exists whenever $E[|1/V|] < \infty$. Here we have $V = \hat{\beta}_2 - \hat{\beta}_1 \sim N(\beta_2 - \beta_1, \eta^2)$. Denote $\mu_2 = \beta_2 - \beta_1$ and $\mathcal{a} = \{v: |v| < \epsilon; \epsilon > 0 \text{ (arb.)}\}$, then

$$\begin{aligned} E[|1/V|] \geq E[(1/|V|)I_{\mathcal{a}}(V)] &= \frac{1}{\eta\sqrt{2\pi}} \int_{-\epsilon}^{\epsilon} (1/|v|)e^{-\frac{1}{2}\left(\frac{v-\mu_2}{\eta}\right)^2} dv \\ &\geq \frac{1}{\eta\sqrt{2\pi}} \int_{0}^{\epsilon} (1/v)e^{-\frac{1}{2}\left(\frac{v-\mu_2}{\eta}\right)^2} dv \quad . \end{aligned}$$

The last integral is improper, and better expressed as

$$\frac{1}{\eta\sqrt{2\pi}} \lim_{\delta \to 0} \int_{\delta}^{\epsilon} (1/v)e^{-\frac{1}{2}\left(\frac{v-\mu_2}{\eta}\right)^2} dv \quad .$$

Now, on the compact set $[\delta, \epsilon]$, the continuous function $\exp\{-\frac{1}{2}(v-\mu_2)^2/\eta^2\}$ attains a minimum value, call it $L(\delta, \epsilon)$ (Goldberg, 1976). On $[\delta, \epsilon]$, it is certainly true, therefore, that $\frac{1}{v}\exp\{-\frac{1}{2}(v-\mu_2)^2/\eta^2\} \geq \frac{1}{v}L(\delta, \epsilon)$. Further, $\exp\{-\frac{1}{2}(v-\mu_2)^2/\eta^2\}$

attains a minimum on $[0,\epsilon]$ — call it $L(\epsilon)$ — and for large enough ϵ, we have $L(\delta,\epsilon) \geq L(\epsilon)$ on $[\delta,\epsilon]$. Given the above, this implies that

$$\int_{\delta}^{\epsilon} (1/v)e^{-\frac{1}{2}\left(\frac{v-\mu_2}{\eta}\right)^2} dv \geq \int_{\delta}^{\epsilon} \frac{L(\epsilon)}{v}\, dv \quad .$$

But $\lim_{\delta\to 0} \int_{\delta}^{\epsilon} \frac{L(\epsilon)}{v} dv$ is infinite, so we can conclude that $E[|1/V|]$ is greater than a positive, infinite value and thus itself diverges. Therefore $E[1/V]$ must diverge as well, i.e.,

$$E[(\hat{\beta}_2 - \hat{\beta}_1)^{-1}] = \infty \quad .$$

But this implies that $E(\hat{J})$ fails to exist. As such, all higher order moments of $\hat{J}$ fail to exist. One obvious consequence of this is that $\mathrm{var}(\hat{J})$ also does not exist.

Now, we noted that in the unusual case of $\hat{J} \notin [x_\tau, x_{\tau+1}]$ the endpoint of this interval which maximizes the likelihood is taken as the ML estimate. This procedure makes it difficult to consider a more general expectation of the join estimator, but by modifying it somewhat and applying Hinkley's (1969) results, the expectation becomes more tractable.

Define, for τ known, a new join estimate as

$$\tilde{J} = \begin{cases} x_\tau & \text{if } \hat{J} \leq x_\tau \\ \hat{J} & \text{if } x_\tau < \hat{J} < x_{\tau+1} \\ x_{\tau+1} & \text{if } \hat{J} \geq x_{\tau+1} \end{cases} \quad . \tag{9}$$

Then in order to calculate $E(\tilde{J})$ from the distribution of $\hat{J}$ we note that $W = \hat{\alpha}_1 - \hat{\alpha}_2$ and $V = \hat{\beta}_2 - \hat{\beta}_1$ have a bivariate Normal distribution with means $\alpha_1 - \alpha_2$ and $\beta_2 - \beta_1$, respectively, variances ν^2 and η^2, defined by (3) and (4), respectively, and correlation

$$\rho = \frac{\operatorname{cov}(W,V)}{\nu\eta} = \left(\frac{\bar{x}_1}{c_1} - \frac{\bar{x}_2}{c_2}\right) \Bigg/ \sqrt{\left(\frac{1}{c_1} + \frac{1}{c_2}\right)\left(\frac{\Sigma_{i=1}^{2} x_i^2}{\tau c_1} + \frac{\Sigma_{i=\tau+1}^{n} x_i^2}{(n-\tau)c_2}\right)} \quad , \tag{10}$$

for $c_1 = \Sigma_{i=1}^{\tau}(x_i - \bar{x}_1)^2$ and $c_2 = \Sigma_{i=\tau+1}^{n}(x_i - \bar{x}_2)^2$. With these we can take

$$\begin{aligned} E(\tilde{J}) &= \int_{-\infty}^{x_\tau} x_\tau f(t)\,dt + \int_{x_\tau}^{x_{\tau+1}} t f(t)\,dt + \int_{x_{\tau+1}}^{\infty} x_{\tau+1} f(t)\,dt \\ &= x_\tau P(\hat{J} < x_\tau) + x_{\tau+1} P(\hat{J} > x_{\tau+1}) + \int_{x_\tau}^{x_{\tau+1}} t f(t)\,dt \qquad (11) \\ &= x_\tau F(x_\tau) + x_{\tau+1}[1 - F(x_{\tau+1})] + \int_{x_\tau}^{x_{\tau+1}} t f(t)\,dt \quad , \end{aligned}$$

where f and F are the p.d.f. and c.d.f., respectively, of $\hat{J}$ (defined in Hinkley, 1969). In our case they are given, using (3), (4) and (10), as

$$f(t) = \frac{bd}{a^3 \eta\nu\sqrt{2\pi}} \left[\Phi\left(\frac{b}{a\sqrt{1-\rho^2}}\right) - \Phi\left(\frac{-b}{a\sqrt{1-\rho^2}}\right)\right] + \frac{\sqrt{1-\rho^2}}{\eta\nu\pi a^2} \exp\left\{\frac{-c}{2(1-\rho^2)}\right\} \quad , \tag{12}$$

and

$$F(t) = L\left\{\frac{\alpha_1-\alpha_2-(\beta_2-\beta_1)t}{a\eta\nu}, \frac{\beta_1-\beta_2}{\eta}, \frac{\eta t-\rho\nu}{a\eta\nu}\right\} + L\left\{\frac{(\beta_2-\beta_1)t-\alpha_1+\alpha_2}{a\eta\nu}, \frac{\beta_2-\beta_1}{\eta}, \frac{\eta t-\rho\nu}{a\eta\nu}\right\} . \tag{13}$$

Here we took Φ as the standard normal c.d.f., L as the standard bivariate normal integral given as

$$L(h,k;g) = \frac{1}{2\pi\sqrt{1-g^2}} \int_h^\infty \int_k^\infty \exp\left\{-\frac{x^2-2gxy+y^2}{2(1-g^2)}\right\} dx\,dy \quad , \tag{14}$$

and

$$a \equiv a(t) = \sqrt{\frac{t^2}{\nu^2} - \frac{2\rho t}{\eta\nu} + \frac{1}{\eta^2}} \quad ,$$

$$b \equiv b(t) = \frac{(\alpha_1-\alpha_2)t}{\nu^2} - \frac{\rho[\alpha_1-\alpha_2+(\beta_2-\beta_1)t]}{\eta\nu} + \frac{\beta_2-\beta_1}{\eta^2} \quad ,$$

$$c \equiv \frac{(\alpha_1-\alpha_2)^2}{\nu^2} - \frac{2\rho(\alpha_1-\alpha_2)(\beta_2-\beta_1)}{\eta\nu} + \frac{(\beta_2-\beta_1)^2}{\eta^2} \quad , \tag{15}$$

$$d \equiv d(t) = \exp\left\{\frac{b^2-ca^2}{2a(1-\rho^2)}\right\} \quad .$$

Tabulated values of (14) are available from the National Bureau of Standards (1959).

In considering evaluation of $E(\widetilde{J})$ from (11), we can apply integration by parts to the integral to find

$$\int_{x_\tau}^{x_{\tau+1}} tf(t)dt = tF(t) \Big|_{x_\tau}^{x_{\tau+1}} - \int_{x_\tau}^{x_{\tau+1}} F(t)dt \quad , \tag{16}$$

so that (11) becomes

$$E(\widetilde{J}) = x_{t+1} - \int_{x_\tau}^{x_{\tau+1}} F(t)dt \quad . \tag{17}$$

Even for relatively simple values of the parameters α_1, α_2, β_1, β_2, η, ν, and ρ, the form of $F(t)$ will be quite complicated, and in order to evaluate the integral in (17) one needs to resort to some sort of numerical quadrature. Still, $F(t)$ is well-defined on the finite interval $[x_\tau, x_{\tau+1}]$, and thus the mean of the sampling estimator $\widetilde{J}$ is indeed finite, and attainable.

The value of $\text{var}(\widetilde{J})$ is also easily attained. As in (11), consider

$$E(\tilde{J}^2) = \int_{-\infty}^{x_\tau} x_\tau 2 f(t)dt + \int_{x_\tau}^{x_{\tau+1}} t^2 f(t)dt + \int_{x_{\tau+1}}^{\infty} x_{\tau+1}^2 f(t)dt$$

$$= x_\tau^2 F(x_\tau) + x_{\tau+1}^2 [1 - F(x_{t+1})] + \int_{x_\tau}^{x_{\tau+1}} t^2 f(t)dt \quad . \tag{18}$$

Again applying integration by parts to the integral yields

$$\int_{x_\tau}^{x_{\tau+1}} t^2 f(t)dt = x_{\tau+1}^2 F(x_{\tau+1}) - x_\tau^2 F(x_\tau) - 2\int^{x_{\tau+1}} tF(t)dt \quad , \tag{19}$$

so that

$$E(\tilde{J}^2) = x_{\tau+1}^2 - 2\int_{x_\tau}^{x_{\tau+1}} tF(t)dt \quad , \tag{20}$$

and thus

$$\mathrm{var}(\tilde{J}) = 2\int_{x_\tau}^{x_{\tau+1}} (x_{t+1} - t)F(t)dt - \left(\int_{x_\tau}^{x_{\tau+1}} F(t)dt\right)^2 \tag{21}$$

(note that once one has computed $E(\tilde{J})$, $\mathrm{var}(\tilde{J})$ involves little additional effort, especially when the slightly more convenient computing form of (21),

$$\mathrm{var}(\tilde{J}) = 2x_{\tau+1}\int_{x_\tau}^{x_{\tau+1}} F(t)dt - \left(\int_{x_\tau}^{x_{\tau+1}} F(t)dt\right)^2 - \int_{x_\tau}^{x_{\tau+1}} tF(t)dt \quad ,$$

is considered). In similar fashion, any mth moment of $\tilde{J}$ can be found as

$$E(\tilde{J}^m) = x_{\tau+1}^m - m\int_{x_\tau}^{x_{\tau+1}} t^{m-1} F(t)dt \quad . \tag{22}$$

For certain values of m, relationships such as (11) or (20) can aid in the computing of (22).

To exemplify this procedure a data set was simulated from the model

$$y_i = \begin{cases} -7.9 + 4.5x_i + e_i & i = 1, \cdots, 7 \\ 8.9 - 7.5x_i + e_i & i = 8, \cdots, 17 \end{cases} \quad , \tag{23}$$

and $e_i \sim$ iid $N(0,4)$. Note that the (true) join occurs at $x = 1.4$. The data are presented in Table I. Applying a result by Hinkley (1971), we find the MLE to be $\hat{J} = 1.3944$, therefore we take $\tilde{J} = 1.3944$. To calculate $E(\tilde{J})$, we apply the data values to (3), (4), and (10) to get $\nu^2 = 4.1659$, $\eta^2 = 1.6542$ and $\rho = 0.3989$. Using these in (13) yields

TABLE I

Simulation Results for Model (23)

x_i	y_i
-.5	-11.0149
-.4	- 8.285
-.2	-10.6704
.4	- 5.1742
.7	- 5.2153
.8	- 8.5362
1.2	- 1.7576
1.6	- 4.1724
3.2	-11.2505
3.7	-21.0943
3.8	-24.9018
4.7	-25.3675
6.1	-37.2594
6.2	-39.3078
6.9	-41.2685
7.0	-42.1461
7.2	-46.5695

Here we have $\tau = 7$, so that $x_\tau = x_7 = 1.2$ and $x_{\tau+1} = x_8 = 1.6$

$$F(t) = L\left(\frac{4.5712t-6.3997}{a(t)},\ 9.3301;\ \frac{0.4899t-0.3101}{a(t)}\right) + L\left(\frac{6.3997-4.5712t}{a(t)},\ -9.3301;\ \frac{0.4899t-0.3101}{a(t)}\right) , \tag{24}$$

with $a(t) = (0.24t^2 - 0.3039t + 0.6045)^{\frac{1}{2}}$, from (15). With these we can evaluate $E(\tilde{J})$ and $\mathrm{var}(\tilde{J})$ from (17) and (21), respectively, using some sort of numerical quadrature. For a series of values of $t_i \in [1.2, 1.6]$, $F(t_i)$ (and $t_i F(t_i)$) can be calculated using the tables in the National Bureau of Standards (1959) publication, or a pre-packaged computer routine of the Cumulative Bivariate Standard Normal Distribution. Adaptive Rhomberg integration (Forsythe et al., 1977) was applied to these values to approximate the integrals in (11) and (21), using a (constant) subinterval length of $h_i = t_{i+1} - t_i = .00025$. The results showed

$$\int_{1.2}^{1.6} F(t)dt \approx 0.19757 \qquad \text{and} \qquad \int_{1.2}^{1.6} tF(t)dt \approx 0.28523 \quad ,$$

so that

$$E(\tilde{J}) \approx 1.40243 \qquad \text{and} \qquad \mathrm{var}(\tilde{J}) \approx 0.02273 \quad .$$

For this example therefore, the results are indeed quite satisfying (recall that $J = 1.4$ and $\tilde{J} = 1.3944$) in terms of the proximity of $E(\tilde{J})$ to J and the relatively small value of $\mathrm{var}(\tilde{J})$.

This is, however, by no means an exhaustive development (e.g., it can be shown that the model structure prohibits $\rho = \pm 1$), and continued investigation of the distribution and sampling behavior of $\tilde{J}$ is underway (Piegorsch, 1982). In general, there are many questions in segmented regression still to be considered; examples include further improvement of point or interval estimates, confidence band construction (cf. Piegorsch, 1981), and smooth line approximations to the segmented model (Watts and Bacon, 1974). The topic is indeed open for future explorations.

References

Blischke, W. R. (1961). Least squares estimators of two intersecting lines. Paper BU-135-M in the Biometrics Unit Mimeo Series, Cornell University, Ithaca, New York.

DeGroot, M. H. (1975). Probability and Statistics. Reading, Massachusetts: Addison-Wesley.

Forsythe, G. E., M. A. Malcolm and C. B. Moler (1977). Computer Methods for Mathematical Computations. Englewood Cliffs: Prentice-Hall.

Goldberg, R. R. (1976). Methods of Real Analysis. New York: Wiley.

Hinkley, D. V. (1969). On the ratio of two correlated normal random variables. Biometrika 56, 635-639.

Hinkley, D. V. (1971). Inference in two-phase regression. J. Amer. Stat. Assoc. 66, 736-743.

National Bureau of Standards (1959). Tables of the Bivariate Normal Distribution Function and Related Functions. Applied Mathematics Series No. 50, Washington, D.C.

Piegorsch, W. W. (1981). A note on confidence bands in bilinear segmented regression. Paper BU-751-M in the Biometrics Unit Series, Cornell University, Ithaca, New York.

Piegorsch, W. W. (1982). A modification of the least squares join point estimator in bilinear segmented regression. M.S. Thesis, Biometrics Unit, Cornell University, Ithaca, New York.

Robison, D. E. (1964). Estimates for the points of intersection of two polynomial regressions. J. Amer. Stat. Assoc. 59, 214-224.

Shabon, S. A. (1980). Change point problem and two-phase regression: An annotated bibliography. Inter. Stat. Review 48, 83-93.

Watts, D. G. and D. W. Bacon (1974). Using an hyperbola as a transition model to fit two-regime straight-line data. Technometrics 16, 369-373.